基于自注意力机制和情绪语义相似性建模的多标签文本情绪分类

倪盈盈

上海交通大学 媒体与传播学院

摘要

【目的】针对现有的多标签文本情绪分类算法难以建模和利用情绪间语义关联性的问题，提出一种基于自注意力机制的情绪关联性建模的多标签文本情绪分类方法。【方法】本文提出基于自注意力机制的情绪关联性建模的多标签文本情绪分类神经网络（AttEmoNet)，通过自注意力情绪关联模块建模情绪标签自身的语义及相似性，并使用基于大规模预训练模型的文本编码器用于编码输入文本为语义向量，最后通过神经网络计算文本语义向量和情绪语义向量的匹配度， 从而实现更加精准的情绪类别识别。【结果】通过在 NLPCC2014 和 GoEmotions 两个公开数据集上的对比实验验证 AttEmoNet 的有效性。结果表明，AttEmoNet 的文本情绪分类性能相比于基线方法 Random、cnsenti、SVM 和 BERT 均有显著提升。相比于现有最优基线方法，AttEmoNet 的分类精准度 Precision 最高提升 13.33%，召回率 Recall 最高提升 21.80%，F1 得分 F1-score 提升 12.74%。同时 AttEmoNet 建模的情绪语义关联矩阵有良好的可解释性，证明其具有较好的建模情绪语义的能力。【局限】大语言模型的出现为多标签文本情绪分类打开了全新的思路，未来将结合 AttEmoNet 与大语言模型各自的优势以实现更加精准、高效的多标签文本情绪分类算法。【结论】本文提出一种基于自注意力机制的情绪关联性建模的多标签文本情绪分类神经网络，提升了文本情绪模型的情绪语义及其关联性建模的能力以及情绪识别的性能，并通过在两个公开数据集上的对比实验验证了研究的有效性。

1 引言

文本情绪分类是数据挖掘研究中的一项重要任务，旨在使用数据挖掘技术（如朴素贝叶斯算法；深度神经网络）从给定文本中发现其情绪倾向，从而有助于对文本内容及其可能影响的深入理解。现如今，随着公共社交媒体平台如微博、推特的兴起，文本情绪分类在舆情研判、社会热点事件追踪等领域已经展现出重要作用。相关工作者使用文本情绪分类算法识别海量社交媒体平台用户发表言论的情绪倾向，从而综合研判舆情走

向以及应对措施。文本情绪分类技术本身，也从传统的简单二分类任务（判断文本情绪倾向是正向或福祥）拓展到多分类任务，即识别文本蕴含的具体情绪，如开心、悲伤、喜欢、愤怒。然而，相比于传统的二分类文本情绪分类，多标签文本情绪分类任务面临着数据稀疏、类别不均衡、情绪语义难以建模等种种挑战。为此，研究者们基于统计学、机器学习和深度学习技术提出了多种多标签文本情绪分类模型。例如，基于情绪字典的文本情绪识别模型[1−3]通过在情绪词典种检索匹配文本，从而确定文本情绪类别。基于朴素贝叶斯和支持向量机的文本情绪词典模型[4]则利用统计学习方法，通过分析与建模词频统计特征从而识别文本情绪概率。随着深度学习在自然语言理解领域的广泛应用[5−7]，以循环神经网络（recurrent neural network, RNN）[8]、大规模预训练模型（pretrained model）[3,9]为代表的深度学习文本情绪识别模型依赖于深度学习在语义表征建模方面的强大能力，在具体文本情绪类别的识别方面取得了显著进展。

然而现阶段的多标签文本情绪分类方法仍有一些重要问题亟待解决。首先是情绪关联性建模。与其他多标签文本分类任务（如新闻类别分类、文本事件分类）不同，多标签文本情绪分类中的类别标签（即情绪）之间存在错综复杂的语义关联性。例如“开心”这一情绪与“喜欢”有着高度的语义重合，直观表现在文本上则为两种情绪常常共现。然而现有的多标签文本情绪分类方法大都将情绪标签视作一般文本分类任务的相互独立标签[10−12,4]，忽视情绪间复杂的语义关联，从而难以深入发掘文本与情绪间的关联性，导致分类算法准确度受限。其次是情绪自身的语义建模。文本情绪分类任务不仅关注文本与情绪的对应关系，同样关注情绪自身蕴含的语义信息。情绪语义信息表征有助于理解情绪的本质以及不同情绪的异同之处，更有助于情绪信息在一些下游任务中（如文本生成）的进一步应用。然而现有的多标签文本情绪分类方法难以建模情绪语义信息表征，从而难以高效挖掘文本情绪分类数据中蕴含的丰富信息。

为解决上述问题，本文基于基本情绪理论[13]和预训练深度学习方法[5,9]，创新性地提出一种用于公共社交平台舆情识别的**预训练情感关联性文本情绪识别模型（AttEmoNet）**。具体来说，AttEmoNet 利用大规模预训练模型 BERT[5]建模文本语义向量，同时使用自注意力机制（self-attention）建模情绪语义特征和情绪关联性。随后，AttEmoNet 使用基于深度学习的多层感知机（multi-layer perceptron, MLP）识别文本情绪概率。接下来通过在中文数据集 NLPCC2014 和英文数据集 GoEmotions 上对比 AttEmoNet 与现有经典文本情绪识别模型的情绪识别能力，以实验验证 AttEmoNet 的情绪识别能力。此外，通过在 NLPCC2014 上的模型参数可视化分析，揭示 AttEmoNet 建

模情绪语义关联性的能力。最后，我们讨论了本文工作的优势与不足，并展望了未来研究方向。

2 相关研究

2.1 基本情绪理论

基本情绪理论由美国心理学家 Paul Ekman 等人于 20 世纪 70 年代提出[13]。该理论认为人类有六种基本情绪，即快乐、悲伤、恐惧、愤怒、惊讶和厌恶。这些基本情绪被认为是跨文化和跨物种的普遍存在。基于情绪基本理论，Ekman 等人[13]通过观察不同文化的人脸表情，发现了一些情绪表情的普遍性。Carroll E. Izard[14]扩展了基本情绪理论，讨论了基本情绪之间的关系以及情绪与认知之间的关系。该研究提出了一种情绪系统的模型，描述了基本情绪之间的关系以及它们是如何相互影响和调节的。例如，作者指出"愤怒"和"厌恶"之间存在一种紧密的关系，而"快乐"和"悲伤"之间则存在着对立的关系。James A. Russell[15]提出了环状情绪理论，对基本情绪理论进行了扩展，强调情绪的构建和主观体验，蕴含情绪间的关联性建模思想。Alan S. Cowen 和 Dacher Keltner[16]探索了人们在自我报告中如何描述和区分不同的情绪体验。该研究发现了相较基本情绪理论更加细粒度的情绪体验，扩展了对情绪的理解，突破了传统基本情绪的观念。它表明情绪是复杂且多样的，可以通过多个离散的情绪类别和连续的渐变来描述和捕捉。总之，基本情绪理论最早提出了情绪的六种基本元素。相关学者基于基本情绪理论，在情绪的构建、情绪之间的关系等方面深入探索，发展出一套逐渐完备的情绪理论框架。

2.2 文本情绪识别

文本情绪识别是一类文本分类任务，旨在根据文本语义特征识别出文本所属的情感类别并分类。根据情绪标签分布的不同，文本情绪识别任务可以分为情绪极性分类（正/负情感，二分类）、情绪类别分类（多类别分类）和情绪标签分类（多标签分类）。文本情绪识别模型包括基于规则的情绪词典方法[1,2,17]、基于统计机器学习的方法[4,18]和基于深度学习的方法[7,3,9,8]。基于规则的情绪词典方法是一种无监督方法，利用情绪词典获取文档中情绪词的情绪值，再通过加权计算确定文档的整体情绪倾向。该方法不考虑词语之间的联系，不考虑词语情绪倾向随上下文内容的变化。常见的情绪词典包括英文词典 General Inquirer、SentiWordNet、Opinion Lexicon、MPQA 等[19]，以及中文词典 HowNet[20]、NTUSD[21]、中文情绪词汇本体库[22]等。基于统计机器学习的方法是一种有监督方法，在带情绪标签的文本数据上训练机器学习分类模型，再将训练好的机器学习分类模型应

用于文本情绪预测任务中。例如谢丽星等[23]提出一种基于支持向量机（support vector machine, SVM）的文本情绪识别模型，将情绪分析过程分为两大策略和四种方法。刘宝芹等[24]提出一种基于朴素贝叶斯（naive Bayesian, NB）的文本情绪识别方法，建立三层树状情绪识别结构。基于深度学习的方法是一种有监督方法，在有情绪标签的文本数据上训练神经网络分类模型，利用神经网络极强的拟合能力来精准预测文本情绪类别。例如苏小英等[25]提出一种基于卷积神经网络的文本情绪识别模型，其中双卷积层结构可以从任意长度语句中抽取特征。孙晓等 [26]提出一种基于深度新年网络的文本情绪识别模型，利用深度信念网络解决了文本特征稀疏的问题。随着大语言模型（large language model, LLM）[27–29]的兴盛，基于预训练的 LLM 方法在文本情绪识别任务中展露头角，并且在大量数据集上取得了最优分类性能。例如，方英兰等[30]使用 BERT 模型获取更完整的文本语义表征，从而更加精准地预测文本情绪类别。马达等[31]在社交网络用户传播他人隐私信息行为的研究问题上对比了多种大语言模型的文本情绪识别能力。

2.3 深度学习与注意力机制

注意力机制（Attention mechanism）最早由 Bahdanau 等[32]提出，是一种深度学习中的用于建模语义序列中不同部分的语义关联性以及关联表征的技术。在自然语言处理中，注意力机制常被用于建模预料上下文之间的语义关联，从而在文本生成、文本分类等任务中实现模型输出结果与上下文的对应。Vaswani 等[5]提出的 Transformer 模型是使用自注意力机制的代表模型。Transformer 模型有着强大的语义表征能力和文本输出能力，是许多文本分类器、文本情感识别方法的基石。

3 基于情绪关联性建模的文本情绪识别方法

现有的文本情绪识别方法难以建模情绪关联性在情绪识别中的重要作用。因此，本文首先提出一种基于情绪关联性建模的文本情绪识别方法（AttEmoNet）。随后，通过在微博文本情绪识别数据集上证明了 AttEmoNet 在文本情绪识别和情绪关联建模方面的优越性。最后，将 AttEmoNet 应用于给定话题下的文本情绪分析中。

3.1 AttEmoNet 方法介绍

本文提出的 AttEmoNet 算法框架如图 1 所示。该框架包含三个模块：文本编码器模块、Attention 文本关联建模模块、情绪识别神经网络模块。其中文本编码器模块使用大规模预训练模型 BERT 将文本输入编码为高维文本语义向量。Attention 文本关联建模模块使用注意力机制，将可训练的情绪本征特征向量变换并输出成蕴含情绪关联性的特征

向量，同时输出**情绪关联性矩阵**。情绪识别神经网络模块将文本语义向量和每个蕴含情绪关联性的情绪特征向量逐个匹配和计算文本包含该情绪的概率。算法最终输出情绪蕴含概率。在训练阶段，使用多标签文本情绪识别数据集训练模型的情绪本征特征向量、Attention 文本关联建模模块以及情绪分类神经网络；在推理阶段，冻结 AttEmoNet 的网络参数，实现端到端的文本情绪识别。

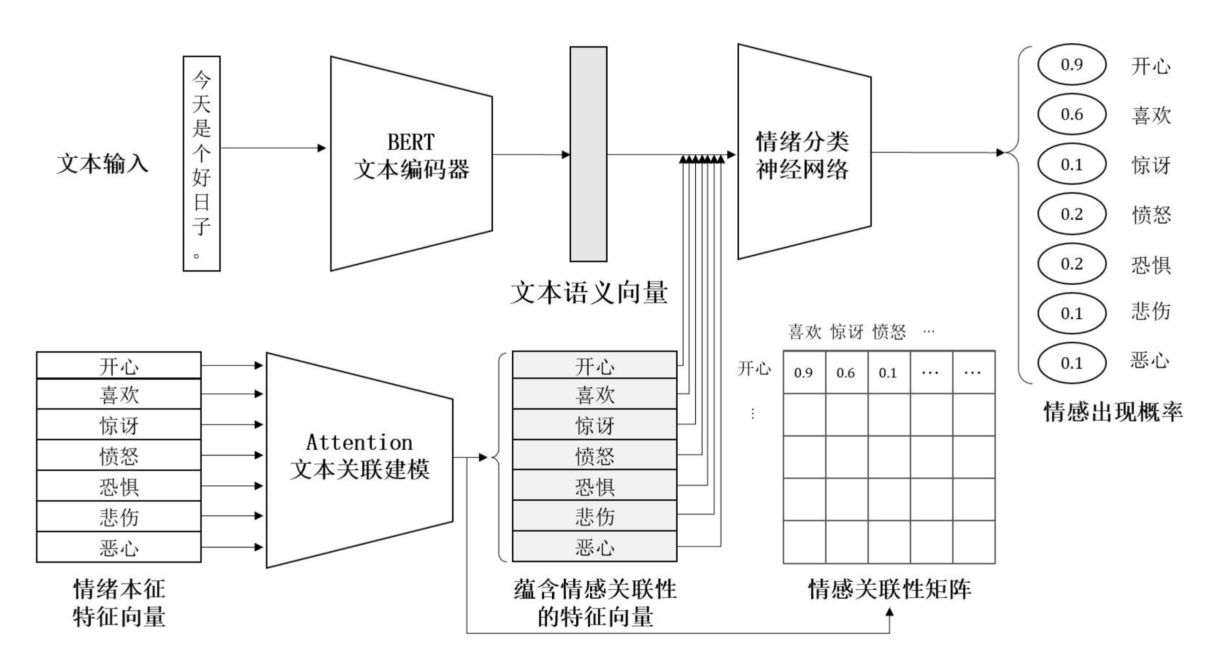


图 1. AttEmoNet 算法框架

3.2 BERT 文本编码器

AttEmoNet 框架中的 BERT 文本编码器，是一个基于 BERT[5]的大型的中文预训练文本编码模型[33]。该模块使用掩码语言模型（masked language model, MLM），从而能生成深度的双向语言表征。BERT 原论文[5]中的实验表明，BERT 在 11 个自然语言处理任务中取得了当时的最优效果，这证明了 BERT 模块在文本语义表征方面的有效性。

形式化地，设原始文本输入为字符序列 $s=\{w_1,w_2,\dots,w_N\}$，则 BERT 的编码过程可以按如式(1)所示形式化表示：

$$v_s^{(senti)}=f_{BERT}(w_1,\dots,w_N)\#(1)$$

其中$v_s^{(senti)}\in\mathbb{R}^D$ 是文本语义表征向量。D 是由 BERT 定义的文本语义表征向量的维度。一般情况下取$D=1{,}024$。

3.3 Attention 文本关联建模模块

Attention 文本关联建模模块使用自注意力机制建模情绪语义的关联性，从而解决现有研究中缺乏对情绪关联性的研究的问题。具体地，自注意力机制采用查询-键-值（query-key-value, QKV）模式。框架中的每一种情绪（如图 1 中的“开心”、“喜欢”、“惊讶”）等都拥有一个可训练的查询向量、键向量、值向量（值向量对应图 1 中的情绪特征本征向量）。首先，对于一种目标情绪（如“开心”），首先取得它的查询向量，使用余弦相似度方法计算该查询向量与其它每一种情绪的键向量的相似度。与每一种其它情绪的相似度即为情绪语义依赖度，即目标情绪的语义表征在多大程度上依赖于该情绪。随后，计算目标情绪的蕴含情绪关联性的特征向量。该向量等于每一种情绪的本征特征向量（值向量）的加权平均，权重即为计算得到的情绪语义依赖度。最后，计算蕴含情绪关联性的特征向量的两两之间的 Pearson 相关系数，输出情绪关联性矩阵。

形式化地，首先使用独热编码来标记每一种情绪（若情绪总数为 K，则第 k 个情绪编码为一个除了第 k 个元素为 1，其他位置元素全为 0 的 K 维向量）。令$S=(s_{jk})_{D\times K}$表示情绪特征本征向量矩阵，$Q=(q_{jk})_{D\times K}$表示情绪查询向量矩阵，s$Z=(z_{jk})_{D\times K}$表示情绪键向量矩阵。则该模块首先取目标情绪 k 的特征本征向量、查询向量和每一种情绪（包括目标情绪自身）的键向量：

$$e_k=S\times x_k,\#(2)$$

$$q_j=Q\times x_j,j=1,2,\ldots,K,\#(3)$$

$$z_j=Z\times x_j,j=1,2,\ldots,K,\#(4)$$

随后，计算目标情绪和每一种情绪的语义依赖相似度：

$$\alpha_{k,j}=softmax\left(\frac{q_k^{\top}z_j}{\sqrt{D}}\right).\#(5)$$

最后计算目标情绪的蕴含情绪关联性的情绪语义向量：

$$e_k^{(att)}=\sum_{j=1}^{K}\alpha_{k,j}\times e_j.\#(6)$$

其中计算结果$e_k^{(att)}$便是蕴含了情绪依赖关系的情绪向量表征，用于在后续步骤中识别文本情绪。

3.4 情绪识别神经网络模块

情绪识别神经网络模块使用一个神经网络计算文本语义表征和情绪语义表征之间的匹配程度，从而预测文本中包含每一种情绪的概率。具体地，给定一句文本的语义表征向量$v_s^{(senti)}$和一种情绪的语义表征向量$e_k^{(att)}$，此模块使用二次形式的神经网络预测文本情绪概率：

$$
\begin{aligned}
\widehat{y}_{s,k} &= sigmoid\left(v_s^{(att)\top} W e_k^{(senti)}\right) \\
&= sigmoid\left(v_s^{(att)\top} (O^\top \Lambda O) e_k^{(senti)}\right) \\
&= sigmoid\left(\left(O v_s^{(att)}\right)^\top \Lambda \left(O e_k^{(senti)}\right)\right)
\end{aligned} \tag{7}
$$

其中$W = O^\top \Lambda O$是情绪识别参数矩阵$W \in \mathbb{R}^{D \times D}$的特征值分解。上述特征值分解变换表明，这一神经网络预测过程等价于将文本语义向量和情绪语义向量经过相同的线性变换后，再逐元素加权平均，权重为W的特征向量。而神经网络的训练过程，相当于优化线性变换和特征向量，使得文本情绪预测概率值接近真实数据标签。

3.5 损失函数

由于 AttEmoNet 面向的文本情绪预测问题是多标签分类问题，因此采用交叉熵作为损失函数。在模型训练过程中，AttEmoNet 的训练目标是最小化损失函数值：

$$
L(\Omega) = -\sum_{i=1}^{N} \sum_{k=1}^{C} y_{i,k} \log\left(p_{i,k}\right)
$$

其中Ω表示 AttEmoNet 中所有的可训练参数，N表示样本数（训练集中的文本数），C表示可能存在的情绪类别数。$y_{i,k}$表示数据集内文本i是否包含情绪k。若$y_{i,k} = 1$则文本i包含情绪k，否则$y_{i,k} = 0$。$p_{i,k}$表示由 AttEmoNet 预测的文本i包含情绪k的概率。

4 文本情绪识别实验

4.1 实验设置

本实验通过在公开微博数据集上对本文提出的文本情绪识别模型 AttEmoNet 和多种文本情绪预测基线模型进行对比测试，验证 AttEmoNet 在文本情绪预测任务上的准确

性以及其建模情绪特征相关性的能力。实验数据集方面，本研究采用 NLPCC2014 和 GoEmotions[34]两个公开数据集。对于 NLPCC2014 数据集，该数据集来源于新浪微博的 45,421 条文本数据，使用人工标注法标注文本情绪标签，包括愤怒、恶心、害怕、开心、喜欢、悲伤、惊讶、无情绪，总共 8 种标签，包含 7 种情绪。每一条文本最多蕴含两种情绪。对于 GoEmotions 数据集，该数据集来源于英文论坛 Reddit 的 58,000 条文本数据，原始数据包含 27 种细粒度情感类别。基于基本情绪理论，我们从中筛选出同 NLPCC2014 数据集的 7 种情绪以及中立情况作为文本情绪识别目标，共筛选出 32,445 条有效样本。接下来，我们分别从每个数据集中按 70%:10%:20%的比例划分训练集、验证集和测试集。

表 1.预处理后的数据集统计数据

数据集	**NLPCC2014**	**GoEmotions**
样本数	45,421	32.445
样本占比（无情绪/中立）	44.1%	56.3%
样本占比（情绪数=1）	38.4%	41.4%
样本占比（情绪数=2）	17.5%	2.2%
样本占比（情绪数=3）	0.0%	0.06%
样本占比（情绪数=4）	0.0%	0.01%

实验环境方面，本文所有模型均使用 Python3.8 实现，深度学习框架为 PyTorch，操作系统为 Linux。实验运行的硬件配置为一台包含两个 2.10GHz Intel Xeon E5-2620 v4 CPU 和一个 NVIDIA Tesla-A100 GPU 的服务器。

4.2 文本情绪预测实验

本文实验主要包含情绪预测实验与情绪特征关联性分析。最后，将 AttEmoNet 情绪预测模型应用于舆情识别中。在文本情绪预测实验，采用的基线模型如下：

- Random：随机预测。对于每一种情绪，以 1/2 的概率将文本归入该类情绪中。一个情绪预测模型的表现是否好过随机预测，是其可用性的基本标准。

- cnsenti：中文情绪分析库(Chinese Sentiment)，基于中国知网情绪词典 HowNet 的情绪预测模型。
- SVM：支持向量机（Support Vector Machine），基于支持向量的情绪预测模型。在实验中，使用 BERT 将文本编码为语义向量，作为 SVM 的输入。
- BERT：基于 Transformer 的预训练大语言模型。使用全连接神经网络（Full-connection Neural Network）作为文本情绪预测任务的下游输出层。

表 4. 文本情绪预测实验结果.

数据集	NLPCC2014			GoEmotions		
模型	Precision ↑ (%)	Recall ↑ (%)	F1-score ↑ (%)	Precision ↑ (%)	Recall ↑ (%)	F1-score ↑ (%)
Random	0.1989	0.4929	0.2636	0.126	0.4894	0.2005
cnsenti	0.1403	0.0943	0.1084	-	-	-
SVM	0.3126	0.5005	0.3247	0.2095	0.4156	0.2786
BERT	0.7215	0.6856	0.6740	0.5546	0.5403	0.5473
AttEmoNet	**0.8177**	**0.7029**	**0.7309**	**0.5808**	**0.6581**	**0.6170**
性能提升	**13.33%**	**2.52%**	**8.44%**	**4.72%**	**21.80%**	**12.74%**

文本情绪预测实验结果如上所示。考虑到多标签分类任务特性，评测指标为微观精度（Micro Precision）、微观召回率（Micro Recall）、微观 F1 得分（Micro F1 Score）。对于上述每个评价指标，得分越高，则模型的文本情绪识别准确度越高。其中由于 GoEmotions 为英文数据集，因此基于中文词典的基线模型 cnsenti 无法识别该数据集的文本情绪。由上述实验结果可知，本文提出的 AttEmoNet 在精度、召回率、F1 得分三个文本情绪预测指标上，均领先于现有文本情绪预测基线模型，其中精度提升最高位 13.33%，召回率最高提升 21.80%，F1 得分最高提升 12.74%。这证明了 AttEmoNet 相比于现有的模型，能够更加准确地预测文本情绪。此外，在基线模型中，BERT+NN 地方法也大幅度优于其他现有方法。其中，BERT+NN 与 cnsenti 的对比，说明了基于 BERT 预训练语言编码的文本情绪预测模型，在微博情绪预测上拥有着比基于规则和情绪词典的传统模型有着更好的性能；BERT+NN 与 BERT+SVM 的对比，说明了基于神经网络的文本情绪预测算法比基于支持向量机（SVM）的算法在微博情绪预测上有着更好的性

能。而相比于基线中的最优模型 BERT+NN，本文提出的 AttEmoNet 方法通过创新性的情绪特征建模模块，进一步提升了基于 BERT 预训练语言编码的文本情绪预测模型的性能。

4.2 可视化实验：情绪特征相关性建模实验

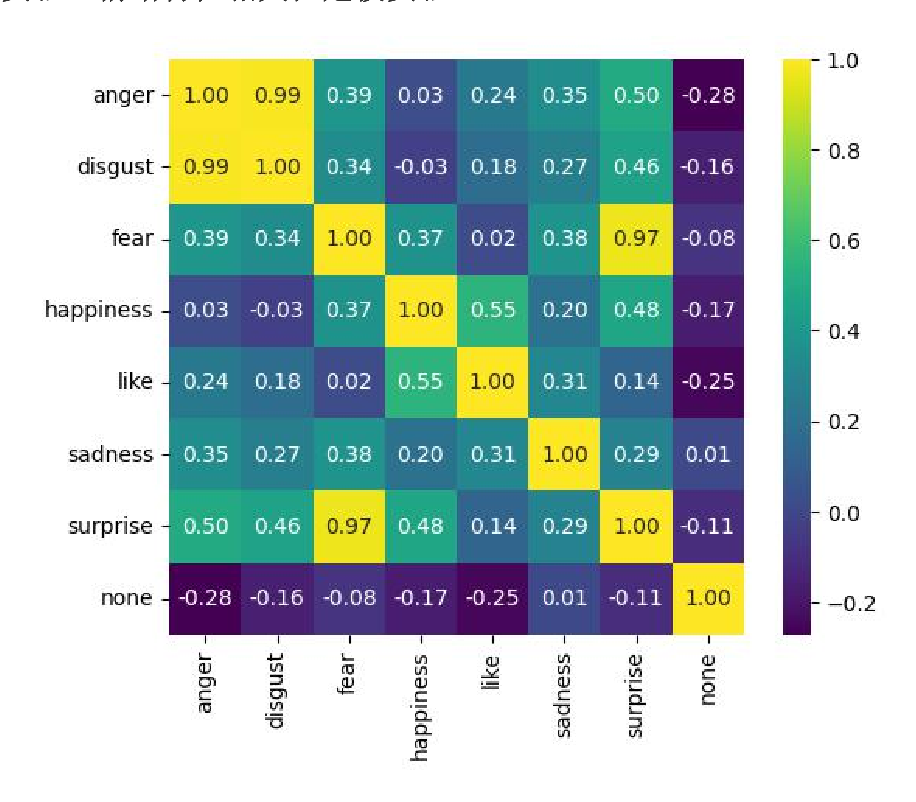


图 2. 情绪特征相关性热力图

这一节以 NLPCC2014 数据集为例，分析 AttEmoNet 建模情绪语义相似性的能力。AttEmoNet 文本情绪预测模型通过 Attention 文本情绪建模模块建模情绪特征之间的关联性，从而提升文本情绪预测精度。本实验阶段主要关注 AttEmoNet 自身的文本情绪特征关联性建模结果。在 AttEmoNet 中，情绪特征使用$e_k^{(att)}$表示，其中 k 表示情绪类别序号。对于任意两种情绪 k1 和 k2，本实验中采用情绪特征的 Pearson 相关系数作为情绪特征相关性度量，记作$Corr(k_1,k_2)$。该相关系数介于-1 与 1 之间。当$Corr(k_1,k_2) > 0$ 时，两种情绪特征之间呈现正相关性（相似性）；当$Corr(k_1,k_2) \approx 0$ 时，两种情绪特征之间呈现不相关性（独立性）；当$Corr(k_1,k_2) < 0$ 时，两种情绪特征之间呈现负相关性（语义相反）。情绪特征相关性计算结果如下图所示，图中包含 7 种情绪：愤怒（anger）、恶心（disgust）、害怕（fear）、开心（happiness）、喜欢（like）、悲伤（sadness）、惊讶（surprise）。图中的每个方块颜色越明亮、相关性值越大，则两种情绪之间关联性

越强。由图 2 可知，与每种情绪关联性最强的三种情绪如下：

- **愤怒**：恶心（0.99）、惊讶（0.50）、害怕（0.39）
- **恶心**：愤怒（0.99）、惊讶（0.46）、害怕（0.34）
- **害怕**：惊讶（0.97）、愤怒（0.39）、悲伤（0.38）
- **开心**：喜欢（0.55）、惊讶（0.48）、害怕（0.37）
- **喜欢**：开心（0.55）、悲伤（0.31）、愤怒（0.24）
- **悲伤**：害怕（0.38）、愤怒（0.35）、喜欢（0.31）
- **惊讶**：害怕（0.97）、愤怒（0.50）、开心（0.48）

从上述结果可见，不同种类的情绪由于其语义不同，或呈现强烈的相关性，或与部分情绪相互独立；部分情绪由于语义具有一致性倾向，往往呈现出较强的聚类特征。例如，“愤怒”与“恶心”同为负面情绪，语义相关性高达 0.99。并且它们同时与“害怕”有着较强的相关性，这说明上述四种情绪在语义内涵方面较为相似，这与人们的直觉相符。与此同时，“开心”与“喜欢”有着较强的相关性，这表明两种直观上的正面情绪在语义内涵方面同样具有相似性。此外，“惊讶”既与“开心”等正面情绪有着较高的语义相似性，也与“害怕”等负面情绪具有较高的语义相似性。这说明“惊讶”作为一种个体由于受到突然变化而感知的情绪，其倾向偏向中立。换言之，“惊讶”既可以与正面情绪共存（如“惊喜”），也可以与负面情绪共存（如“惊恐”）。

6 结论

网络社交平台极易出现大规模争议性网络议题，甚至不少议题极易发酵为网络舆情并进入大规模情绪化的非理性传播。现有的情绪识别模型难以建模情绪关联性，情绪预测精度有待提高。为解决上述问题，本研究首先进行广泛深入的相关研究调研，基于基本情绪理论和深度学习技术，创新性地提出一种大规模预训练文本情绪识别方法（AttEmoNet），用于实现精准的网络社交平台文本情绪识别与情绪关联性建模，并通过在真实文本情绪识别中文数据集 NLPCC2014 和英文数据集 GoEmotions 上的大规模对比实验验证了 AttEmoNet 的精准文本情绪识别能力。情绪识别对比实验表明，AttEmoNet 在文本情绪识别的精度（Precision）、召回率（Recall）、F1 得分（F1 Score）三个指标上相比于最优基线方法 BERT 分别提升了 13.33%、2.52%和 8.44%，有效提升了文本情绪识别的精准度。情绪特征相关性实验表明，具有相似情绪色彩（正面/负面）的情绪具有较强的语义相关性；“惊讶”情绪与正面情绪和负面情绪都具有较高的语义相关性，在

情绪相关性图中成为连接二者的桥梁。

7 讨论

本文的研究意义在于：首先，本文在理论层面将基本情绪理论与深度学习技术有机结合，创新性地提出一种大规模与训练文本情绪识别方法（AttEmoNet），并通过真实数据集上的大规模实验验证了该方法的准确识别文本情绪和建模情绪语义相关性的能力。在文本情绪识别这一任务中，情绪识别效果的准确性是相关研究的核心议题，也是公众舆情监督的重要技术保证。因此，AttEmoNet 在实验中体现的高性能，对于增强公众舆情监督效果无疑具有重要意义。其次，本文在实践层面首次探究了“女性主义”这一高敏感度微博话题下的公众情绪分布及随时间演变规律。此外，借助 AttEmoNet 的情绪语义相关性建模能力，本文同时分析了该话题下不同情绪之间的相关性关系，为相关的舆情监督给出了重要数据参考。

同时，本研究仍存在一些不足。首先，受到可获取数据量限制，AttEmoNet 所构建的训练语料库仍不足以发挥模型最大性能，在未来研究中需要进一步提升文本数据量。其次，在文本语义解析能力方面，AttEmoNet 方法对于反话、讽刺等隐含信息量较大的文本情绪识别效果仍然有待提升。在未来的研究计划中，一方面我们可以通过扩大数据集、优化模型架构等方式进一步提升文本情绪识别能力；另一方面，随着大语言模型（Large Language Model, LLM）的兴起（如 ChatGPT），我们可以结合大语言模型在文本生成、涌现能力的优势，以及 AttEmoNet 在语义建模强、计算成本低的优势，设计更加高效的文本情绪识别模型。此外，在线社交平台的议题、用户分布错综复杂，蕴含信息丰富。如何利用丰富的议题与用户信息助力文本情绪识别与舆情监督、探索情绪识别和情绪语义建模结果的下游应用，我们亦认为是一个重要的未来研究方向。

参考文献

[1] 王志涛, 於志文, 郭斌, 等. 基于词典和规则集的中文微博情感分析[J]. 计算机工程与应用, 2015, 51(8): 218-225.

[2] 董丽丽, 赵繁荣, 张翔. 基于领域本体、情感词典的商品评论倾向性分析[J]. 计算机应用与软件, 2014, 31(12): 104-108+194.

[3] 陈珂, 谢博, 朱兴统. 基于情感词典和 Transformer 模型的情感分析算法研究[J/OL]. 南京邮电大学学报(自然科学版), 2020, 40(1): 55-62. DOI:10.14132/j.cnki.1673-5439.2020.01.009.

[4] NEETHU M S, RAJASREE R. Sentiment analysis in twitter using machine learning techniques[C/OL]//2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). 2013: 1-5. DOI:10.1109/ICCCNT.2013.6726818.

[5] VASWANI A, SHAZEER N, PARMAR N, 等. Attention Is All You Need[M/OL]. arXiv, 2023[2023-08-28]. http://arxiv.org/abs/1706.03762. DOI:10.48550/arXiv.1706.03762.

[6] MIKOLOV T, SUTSKEVER I, CHEN K, 等. Distributed Representations of Words and Phrases and their Compositionality[EB/OL]//arXiv.org. (2013-10-16)[2023-08-28]. https://arxiv.org/abs/1310.4546v1.

[7] YADAV A, VISHWAKARMA D K. Sentiment analysis using deep learning architectures: a review[J/OL]. Artificial Intelligence Review, 2020, 53(6): 4335-4385. DOI:10.1007/s10462-019-09794-5.

[8] ZHANG T, ZHENG W, CUI Z, 等. Spatial-Temporal Recurrent Neural Network for Emotion Recognition[J/OL]. IEEE Transactions on Cybernetics, 2019, 49(3): 839-847. DOI:10.1109/TCYB.2017.2788081.

[9] CORTIZ D. Exploring Transformers in Emotion Recognition: a comparison of BERT, DistillBERT, RoBERTa, XLNet and ELECTRA[J].

[10] 钟佳娃, 刘巍, 王思丽, 等. 文本情感分析方法及应用综述[J]. 数据分析与知识发现, 2021, 5(6): 1-13.

[11] 李然, 林政, 林海伦, 等. 文本情绪分析综述[J]. 计算机研究与发展, 2018, 55(1): 30-52.

[12] 王希雅, 张宁, 程馨. 文本细粒度情绪识别方法与应用综述[J]. 计算机科学, 2023, 50(S1): 590-596.

[13] EKMAN P. Emotion in the human face: guide-lines for research and an integration of findings[M/OL]. New York, Pergamon Press, 1972[2023-08-29]. http://archive.org/details/emotioninthehuma0000unse.

[14] IZARD C E. Human Emotions[M/OL]. Boston, MA: Springer US, 1977[2023-08-29]. http://link.springer.com/10.1007/978-1-4899-2209-0. DOI:10.1007/978-1-4899-2209-0.

[15] RUSSELL J. A Circumplex Model of Affect[J/OL]. Journal of Personality and Social

Psychology, 1980, 39: 1161-1178. DOI:10.1037/h0077714.
[16]COWEN A S, KELTNER D. Self-report captures 27 distinct categories of emotion bridged by continuous gradients[J/OL]. Proceedings of the National Academy of Sciences, 2017, 114(38): E7900-E7909. DOI:10.1073/pnas.1702247114.
[17]XU L, LI L, JIANG Z, 等. A Novel Emotion Lexicon for Chinese Emotional Expression Analysis on Weibo: Using Grounded Theory and Semi-Automatic Methods[J/OL]. IEEE Access, 2021, 9: 92757-92768. DOI:10.1109/ACCESS.2020.3009292.
[18]RASTOGI A, SINGH R, ATHER D. Sentiment Analysis Methods and Applications–A Review[C/OL]//2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART). 2021: 391-395. DOI:10.1109/SMART52563.2021.9676260.
[19]DENG S, SINHA A P, ZHAO H. Adapting sentiment lexicons to domain-specific social media texts[J/OL]. Decision Support Systems, 2016: S016792361630183X. DOI:10.1016/j.dss.2016.11.001.
[20]FU X, LIU W, XU Y, 等. Combine HowNet Lexicon to Train Phrase Recursive Autoencoder for Sentence-Level Sentiment Analysis[J/OL]. Neurocomputing, 2017, 241(JUN.7): 18-27. DOI:10.1016/j.neucom.2017.01.079.
[21]吴杰胜, 陆奎. 基于多部情感词典和规则集的中文微博情感分析研究[J]. 计算机应用与软件, 2019, 36(9): 93-99.
[22]胡召亚, 张顺香. 基于关键句提取的中文微博情感计算[J/OL]. 阜阳师范学院学报(自然科学版), 2019, 36(3): 92-96. DOI:10.14096/j.cnki.cn34-1069/n/1004-4329(2019)03-092-05.
[23]谢丽星, 周明, 孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取[J]. 中文信息学报, 2012, 26(1): 73-83.
[24]刘宝芹, 牛耘. 多层次中文微博情绪分析[J]. 计算机技术与发展, 2015, 25(11): 23-26.
[25]苏小英, 孟环建. 基于神经网络的微博情感分析[J]. 计算机技术与发展, 2015, 25(12): 161-164+168.
[26]孙晓, 彭晓琪, 胡敏, 等. 基于多维扩展特征与深度学习的微博短文本情感分析[J]. 电子与信息学报, 2017, 39(9): 2048-2055.
[27]ZHAO W X, ZHOU K, LI J, 等. A Survey of Large Language Models[M/OL]. arXiv, 2023[2023-08-29]. http://arxiv.org/abs/2303.18223. DOI:10.48550/arXiv.2303.18223.
[28]徐继敏, 严若冰. 论大语言模型材料的证据属性——以 ChatGPT 和文心一言为例[J/OL]. 四川师范大学学报(社会科学版): 1-15. DOI:10.13734/j.cnki.1000-5315.2023.0316.
[29]严豫, 杨笛, 尹德春. 融合大语言模型知识的对比提示情感分析方法[J]. 情报杂志: 1-9.
[30]方英兰, 孙吉祥, 韩兵. 基于 BERT 的文本情感分析方法的研究[J]. 信息技术与信息化,

2020(2): 108-111.

[31] 马达, 卢嘉蓉, 朱侯. 基于文本情绪分类的社交网络用户传播他人隐私信息行为研究[J/OL]. 情报科学, 2023, 41(2): 60-68. DOI:10.13833/j.issn.1007-7634.2023.02.008.

[32] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[M/OL]. arXiv, 2016[2023-08-29]. http://arxiv.org/abs/1409.0473. DOI:10.48550/arXiv.1409.0473.

[33] CUI Y, CHE W, LIU T, 等. Pre-Training With Whole Word Masking for Chinese BERT[J/OL]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 3504-3514. DOI:10.1109/TASLP.2021.3124365.

[34] DEMSZKY D, MOVSHOVITZ-ATTIAS D, KO J, 等. GoEmotions: A Dataset of Fine-Grained Emotions[C/OL]//JURAFSKY D, CHAI J, SCHLUTER N, 等. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020: 4040-4054[2024-04-21]. https://aclanthology.org/2020.acl-main.372. DOI:10.18653/v1/2020.acl-main.372.